MITRE NIST Challenge Update

Jim Houchins (lead)
Jeff Gold
Shreeja Kikkisetti (intern)
Mark Krangle
Nick Maynard

August, 2020.

Approved for Public Release; Distribution Unlimited.
Public Release Case Number 20-2337

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

Agenda

- Introduction
- Datasets
- Data Preparation
- Algorithms/Analysis Approaches
- Lessons learned

Who is MITRE

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

- As a not-for-profit organization, MITRE works in the public interest across federal, state and local governments, as well as industry and academia. We bring innovative ideas into existence in areas as varied as artificial intelligence, intuitive data science, quantum information science, health informatics, space security, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience.

- We operate FFRDCs—federally funded research and development centers. FFRDCs are unique organizations that assist the United States government with scientific research and analysis; development and acquisition; and systems engineering and integration. We also have an independent research program that explores new and expanded uses of technologies to solve our sponsors' problems.

Above text from: https://www.mitre.org/about/corporate-overview

Previous Research Found In This Space

- Improving BLE Distance Estimation and Classification Using TX Power and Machine Learning: A Comparative Analysis
  - M. Al Qathrady, A. Helmy, MSWiM ’17, November 21–25, 2017, Miami, FL, USA
- A Comprehensive Study of Bluetooth Signal Parameters for Localization
  - A Hossain, W. Soh, 2007 IEEE, The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications
- Inferring distance from Bluetooth signal strength: a deep dive
  - P. Dehaye, 2020. https://medium.com/personaldata-io/inferring-distance-from-bluetooth-signal-strength-a-deep-dive-fe7badc2bb6d
- Extended Gradient Predictor and Filter for Smoothing RSSI
  - F. Subhan, S. Ahmed, et al. 2014. 16th International Conference on Advanced Communication Technology

Datasets

MITRE Range Angled Structured Set

- 2 phone users in a room with iPhones
- Users have phone in various locations on body (see figure 1)
- Users rotate every 15 seconds.
- Varies in distance from 3-15 feet.

During competition, we had 74 sets to use

Protocol for collection found here

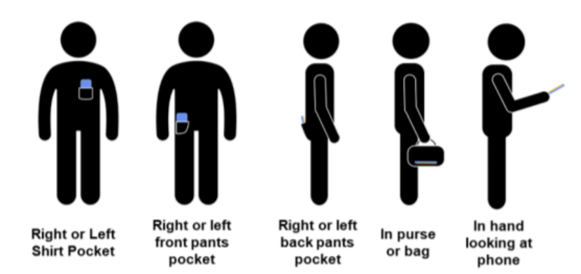

Right or Left Shirt Pocket
Right or left front pants pocket
Right or left back pants pocket
In purse or bag
In hand looking at phone

Figure 1

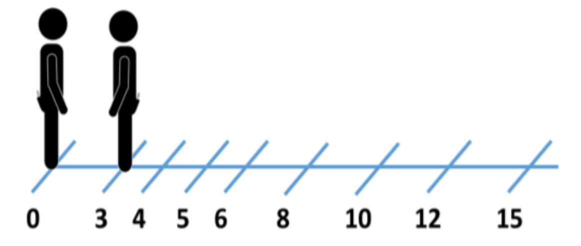
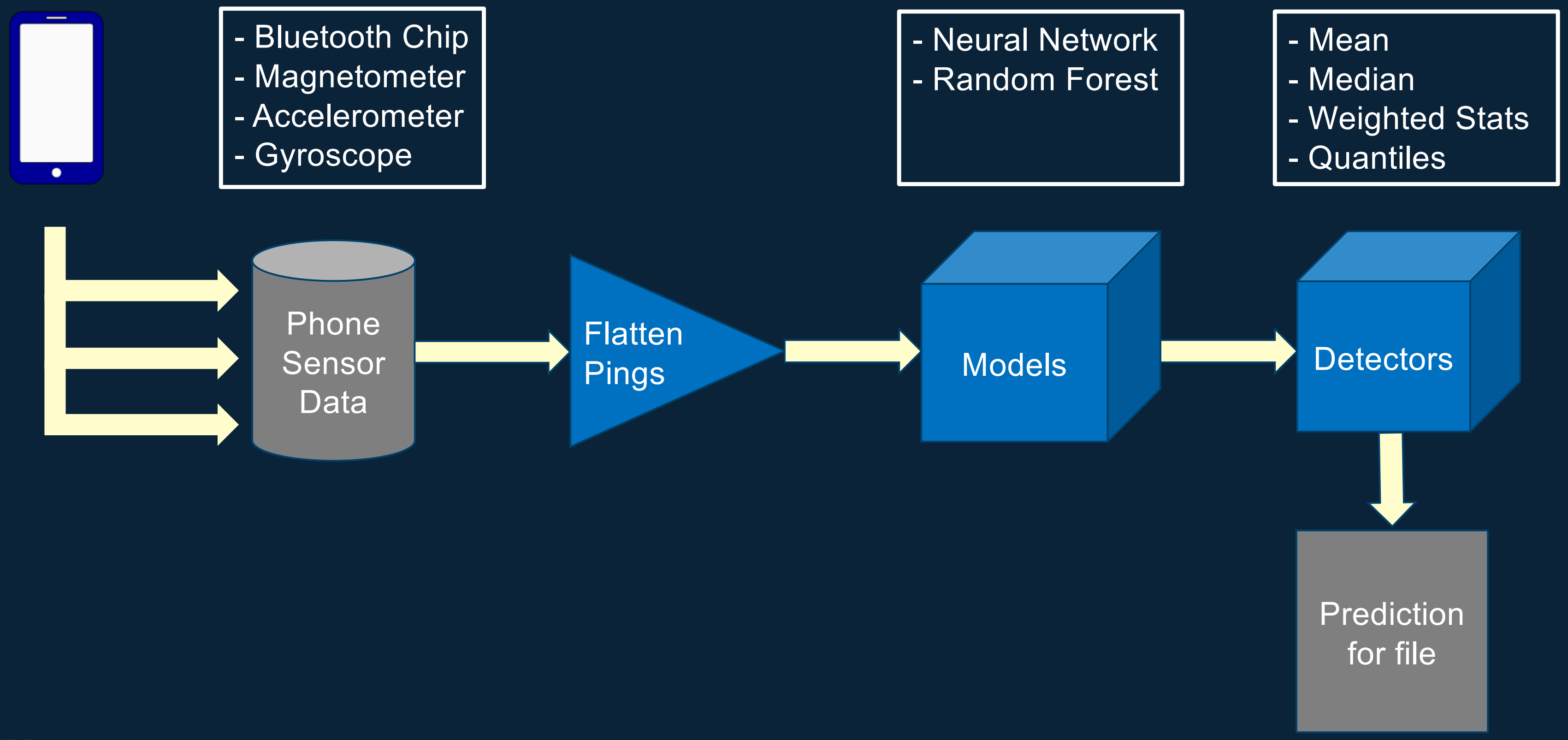


Figure 2

Pipeline Workflow



- Bluetooth Chip
- Magnetometer
- Accelerometer
- Gyroscope

- Neural Network
- Random Forest

- Mean
- Median
- Weighted Stats
- Quantiles

Phone Sensor Data

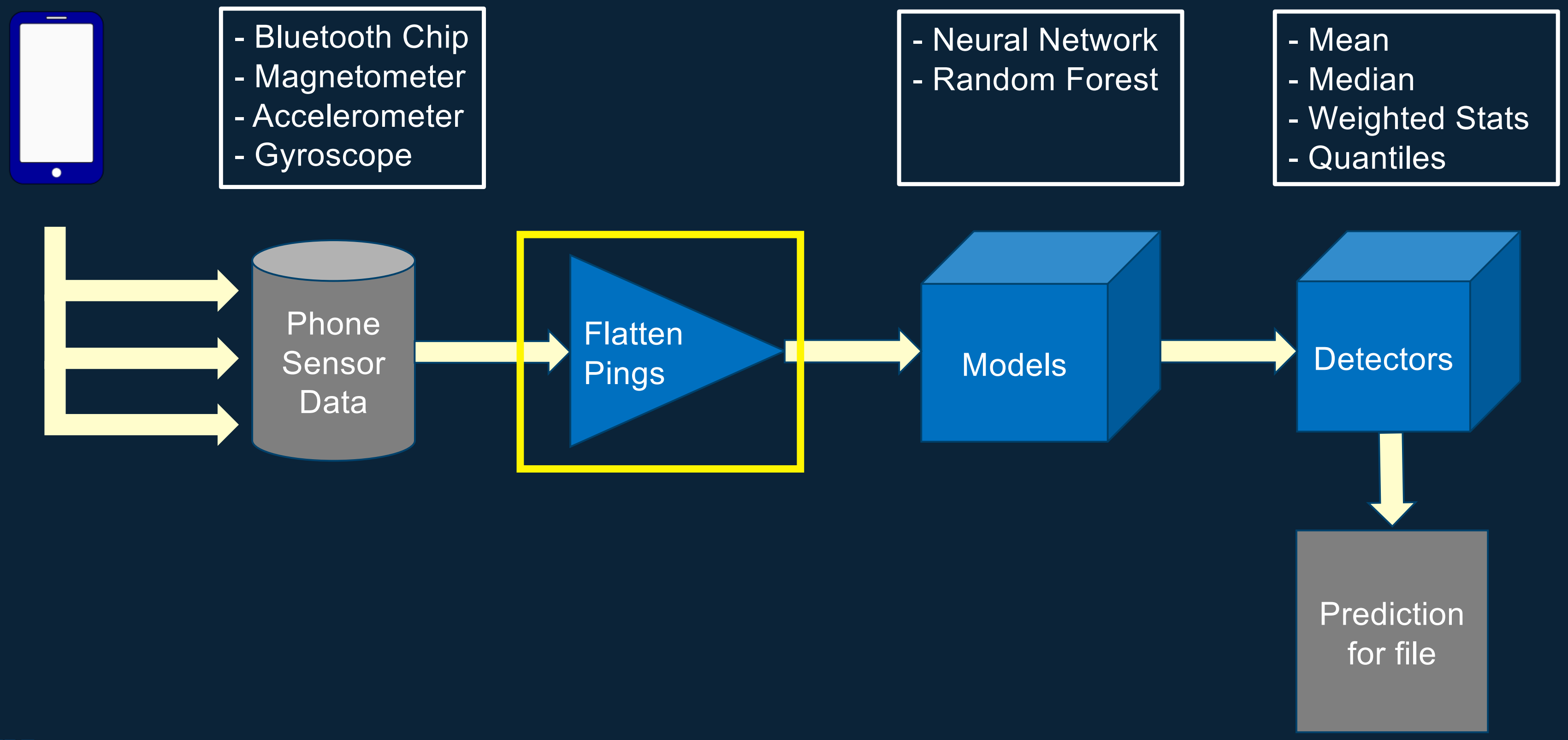
Flatten Pings

Models

Detectors

Prediction for file

Pipeline Workflow



- Bluetooth Chip
- Magnetometer
- Accelerometer
- Gyroscope

- Neural Network
- Random Forest

- Mean
- Median
- Weighted Stats
- Quantiles

Phone Sensor Data

Flatten Pings

Models

Detectors

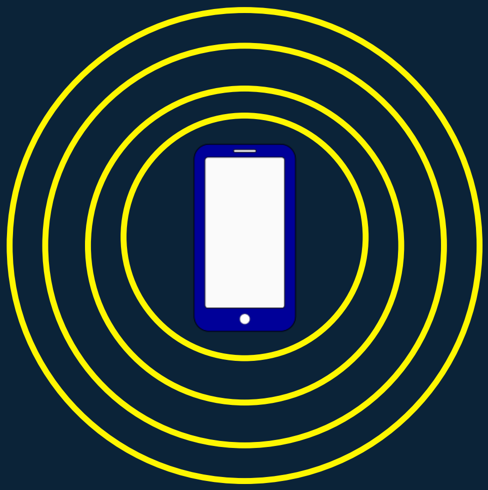
Prediction for file

Flattening Pings

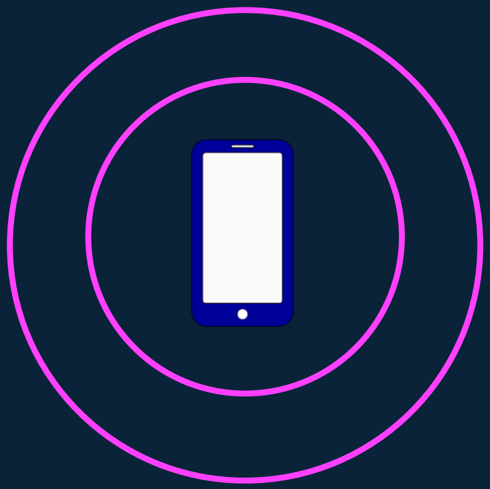
Problem: Sensors ping at different rates, but a symmetric data input is needed to train our model

Solution: Create frame that keeps current value and then updates with new data.

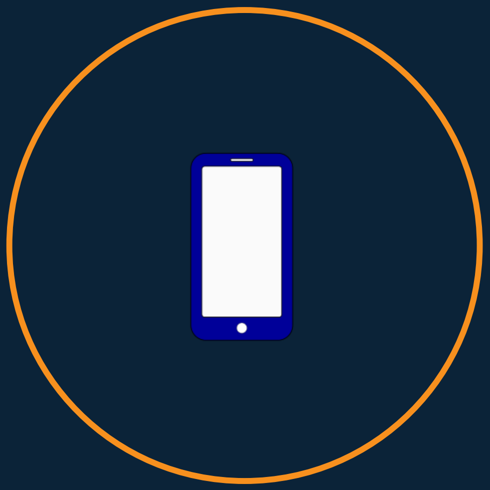
Bluetooth RSSI



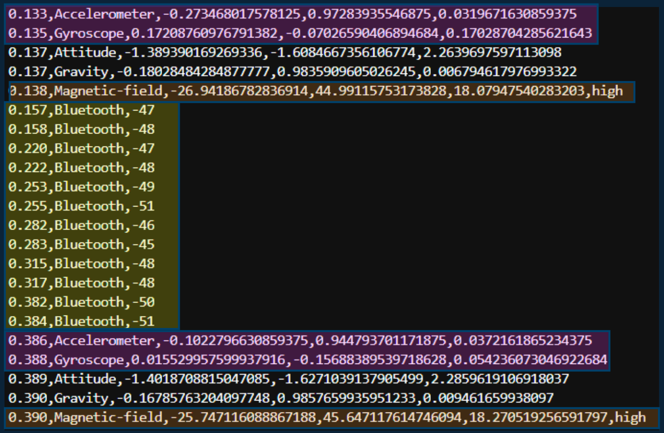
Accelerometer/Gyroscope



Magnetometer

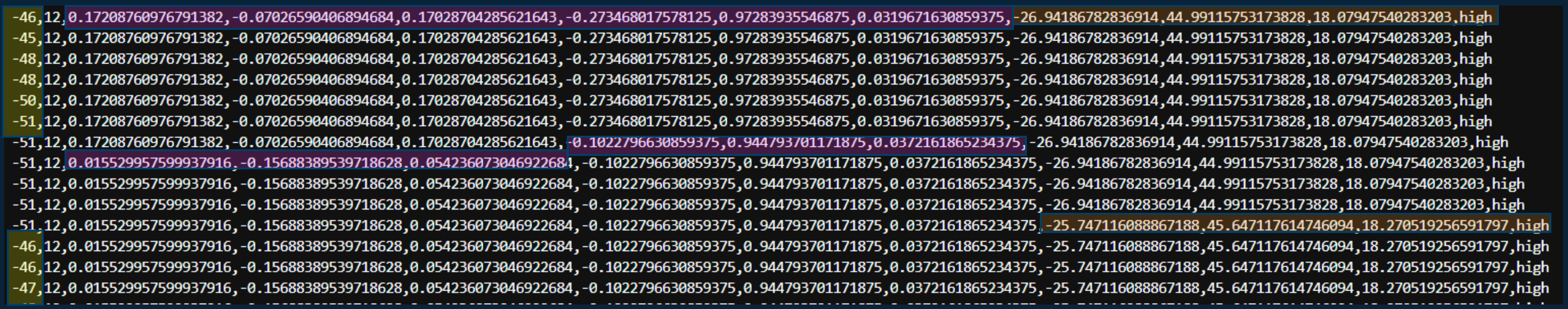


Flattening Pings pt. 2



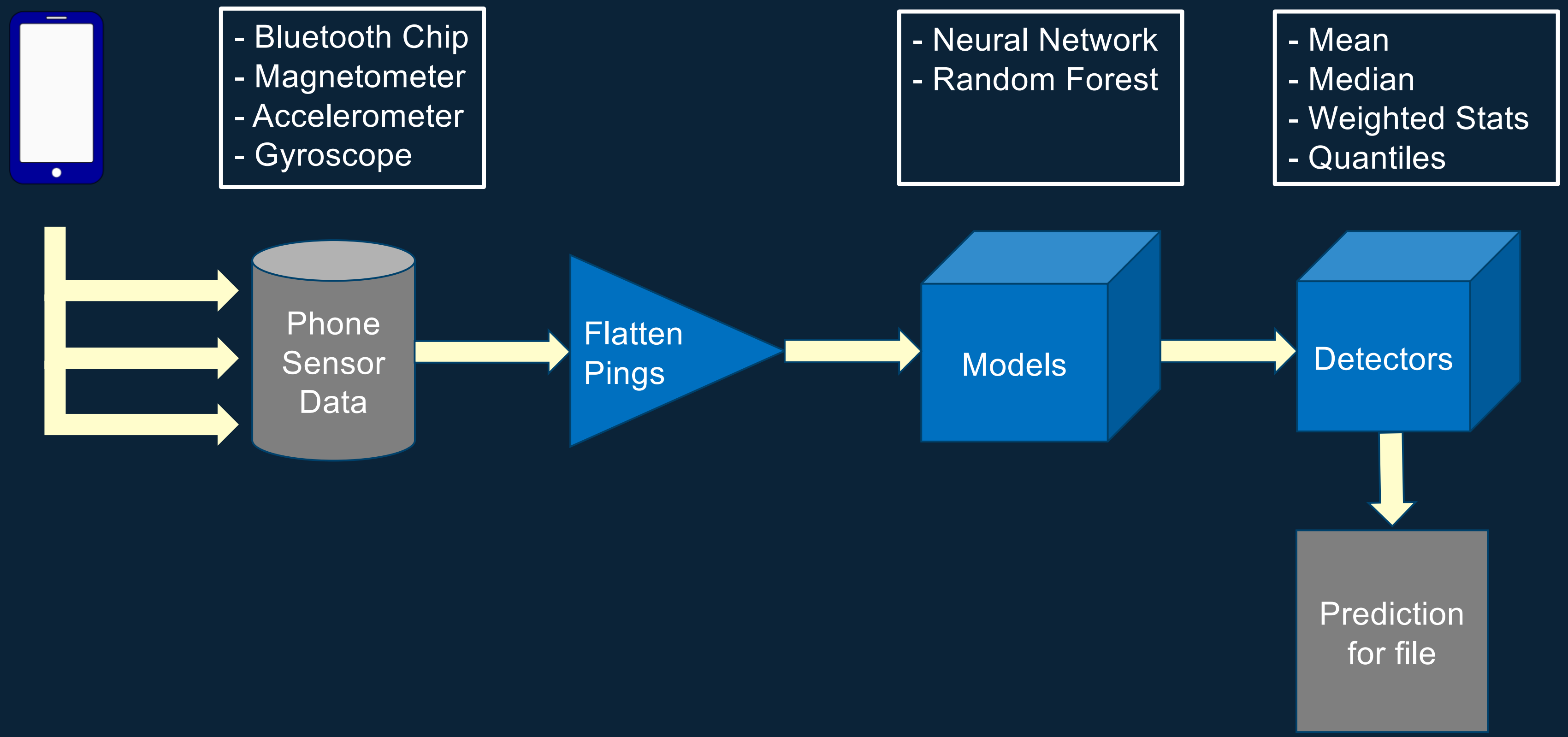
```
0.133,Accelerometer,-0.273468017578125,0.97283935546875,0.0319671630859375
0.135,Gyroscope,0.17208760976791382,-0.07026590406894684,0.17028704285621643
0.137,Attitude,-1.389390169269336,-1.6084667356106774,2.2639697597113098
0.137,Gravity,-0.18028484284877777,0.9835909605026245,0.006794617976993322
0.138,Magnetic-field,-26.94186782836914,44.99115753173828,18.07947540283203,high
0.157,Bluetooth,-47
0.158,Bluetooth,-48
0.220,Bluetooth,-47
0.222,Bluetooth,-48
0.253,Bluetooth,-49
0.255,Bluetooth,-51
0.282,Bluetooth,-46
0.283,Bluetooth,-45
0.315,Bluetooth,-48
0.317,Bluetooth,-48
0.382,Bluetooth,-50
0.384,Bluetooth,-51
0.386,Accelerometer,-0.1022796630859375,0.944793701171875,0.0372161865234375
0.388,Gyroscope,0.015529957599937916,-0.15688389539718628,0.054236073046922684
0.389,Attitude,-1.4018708815047085,-1.6271039137905499,2.2859619106918037
0.390,Gravity,-0.16785763204097748,0.9857659935951233,0.009461659938097
0.390,Magnetic-field,-25.747116088867188,45.647117614746094,18.270519256591797,high
```

- Column values update at every new sensor ping
- Value in column is the most recent value of the sensor.
- After reading in this way null values are dropped.



```
-46,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-45,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-48,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-48,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-50,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
-46,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
-46,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
-47,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
```

Pipeline Workflow



- Bluetooth Chip
- Magnetometer
- Accelerometer
- Gyroscope

Phone Sensor Data

Flatten Pings

- Neural Network
- Random Forest

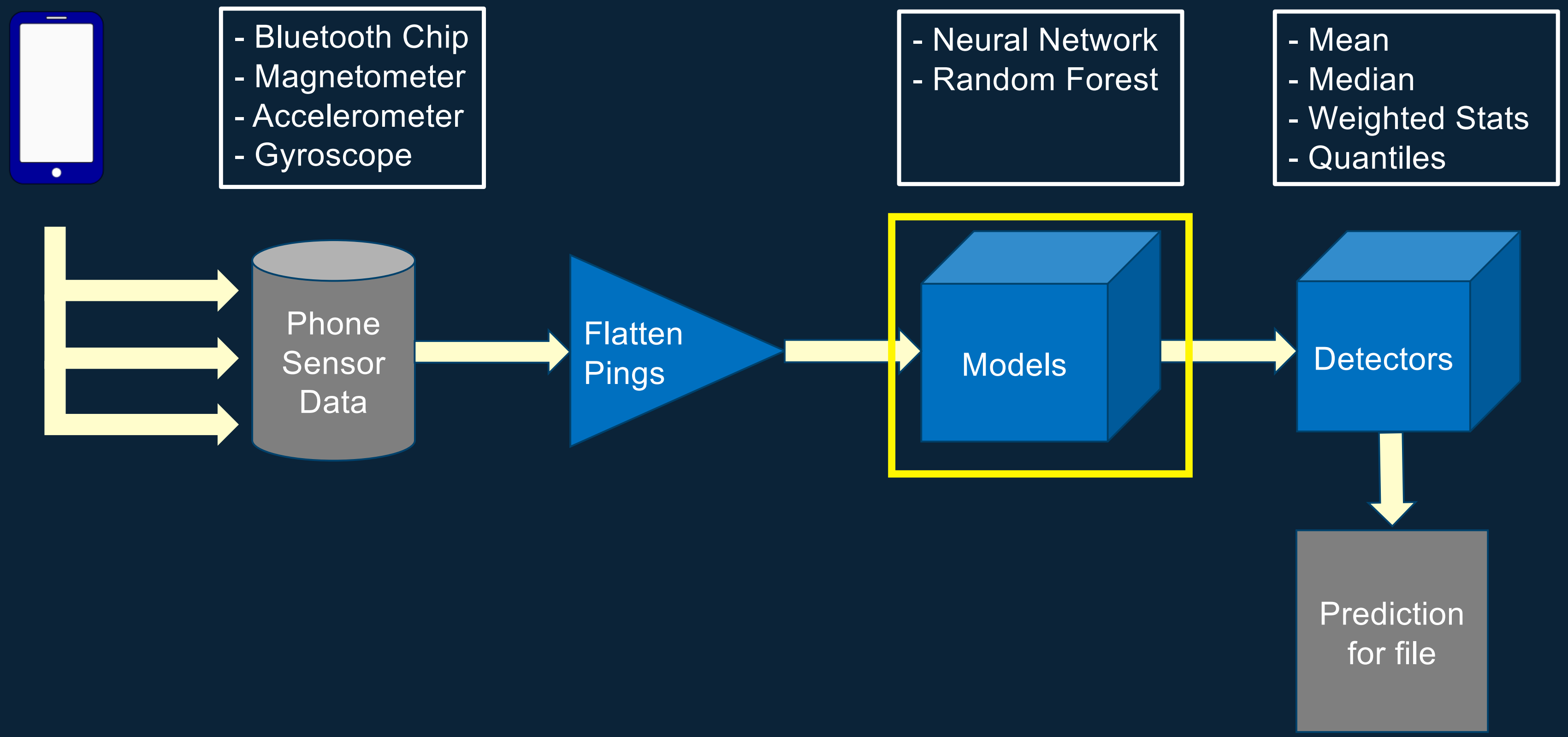
Models

- Mean
- Median
- Weighted Stats
- Quantiles

Detectors

Prediction for file

Pipeline Workflow



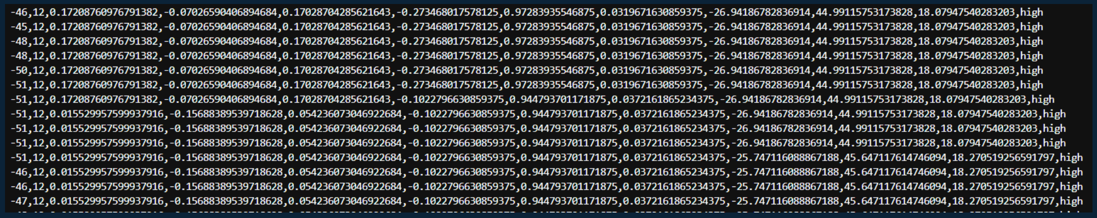
- Bluetooth Chip
- Magnetometer
- Accelerometer
- Gyroscope

- Neural Network
- Random Forest

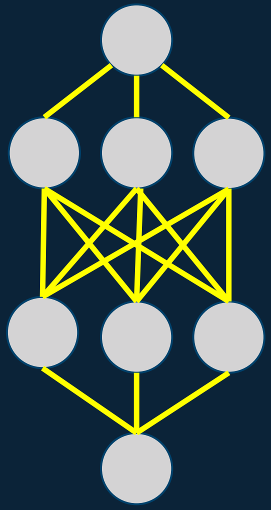
- Mean
- Median
- Weighted Stats
- Quantiles

Sensor Data to Distance Prediction

Input Data

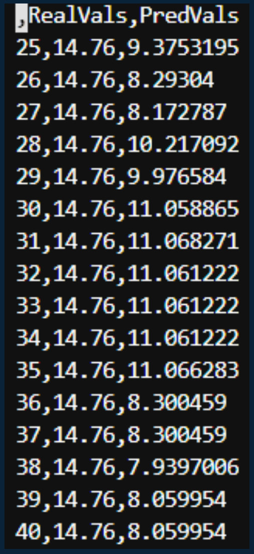


```
-46,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-45,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-48,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-48,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-50,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.273468017578125,0.97283935546875,0.0319671630859375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.17208760976791382,-0.07026590406894684,0.17028704285621643,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-26.94186782836914,44.99115753173828,18.07947540283203,high
-51,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
-46,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
-46,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
-47,12,0.015529957599937916,-0.15688389539718628,0.054236073046922684,-0.1022796630859375,0.944793701171875,0.0372161865234375,-25.747116088867188,45.647117614746094,18.270519256591797,high
```



Model

Model Output



```
,RealVals,PredVals
25,14.76,9.3753195
26,14.76,8.29304
27,14.76,8.172787
28,14.76,10.217092
29,14.76,9.976584
30,14.76,11.058865
31,14.76,11.068271
32,14.76,11.061222
33,14.76,11.061222
34,14.76,11.061222
35,14.76,11.066283
36,14.76,8.300459
37,14.76,8.300459
38,14.76,7.9397006
39,14.76,8.059954
40,14.76,8.059954
```

Neural Network Model

Model Format

Input Layer

- Varied inputs from just the Bluetooth RSSI to the Bluetooth RSSI, Magnetometer, Accelerometer.

Hidden Layers

- Varied from 1-10

Output Layer

- Linear

Loss Function:

- Mean Squared Error
- Mean Average Error

Training Parameters

Batch size

- Varied from 1-2000

Learning rate

- Varied from 0.001-0.01

Epochs

- Varied from 5 – 150

Train/Test split

- Varied from 5/95 – 99/1

Random Forest Model

Model Format

Criteria
- Mean Squared Error

Search Methods:
- Random Search
- Grid Search

Training Parameters

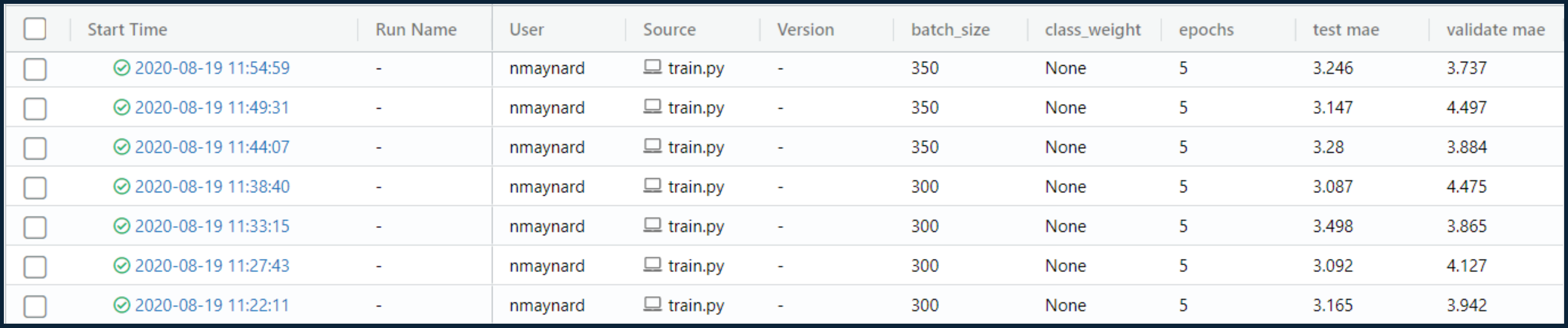
Max Leaf Depth
- Varied from 1-2000

Trees
- Varied from 0.001-0.01

Train/Test split
- Varied from 5/95 – 99/1

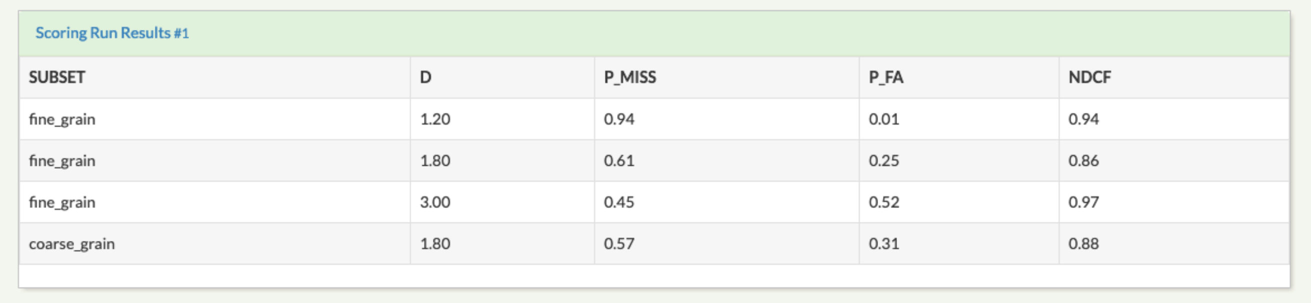
Rapid Iteration Through Models mlflow

- Used mlflow with shell scripts to iterate through multiple models
- Once pipeline was constructed could iterate through multiple models
- Pipeline not complete till last week of competition
- Went from single digit runs per day manually to over a hundred, and that number was limited only by hardware



□	Start Time	Run Name	User	Source	Version	batch_size	class_weight	epochs	test mae	validate mae
□	✓ 2020-08-19 11:54:59	-	nmaynard	train.py	-	350	None	5	3.246	3.737
□	✓ 2020-08-19 11:49:31	-	nmaynard	train.py	-	350	None	5	3.147	4.497
□	✓ 2020-08-19 11:44:07	-	nmaynard	train.py	-	350	None	5	3.28	3.884
□	✓ 2020-08-19 11:38:40	-	nmaynard	train.py	-	300	None	5	3.087	4.475
□	✓ 2020-08-19 11:33:15	-	nmaynard	train.py	-	300	None	5	3.498	3.865
□	✓ 2020-08-19 11:27:43	-	nmaynard	train.py	-	300	None	5	3.092	4.127
□	✓ 2020-08-19 11:22:11	-	nmaynard	train.py	-	300	None	5	3.165	3.942

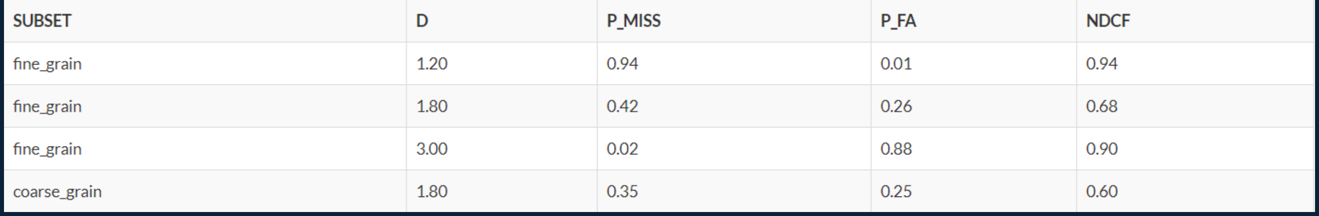
Rapid Iteration Through Models mlflow



Scoring Run Results #1

SUBSET	D	P_MISS	P_FA	NDCF
fine_grain	1.20	0.94	0.01	0.94
fine_grain	1.80	0.61	0.25	0.86
fine_grain	3.00	0.45	0.52	0.97
coarse_grain	1.80	0.57	0.31	0.88

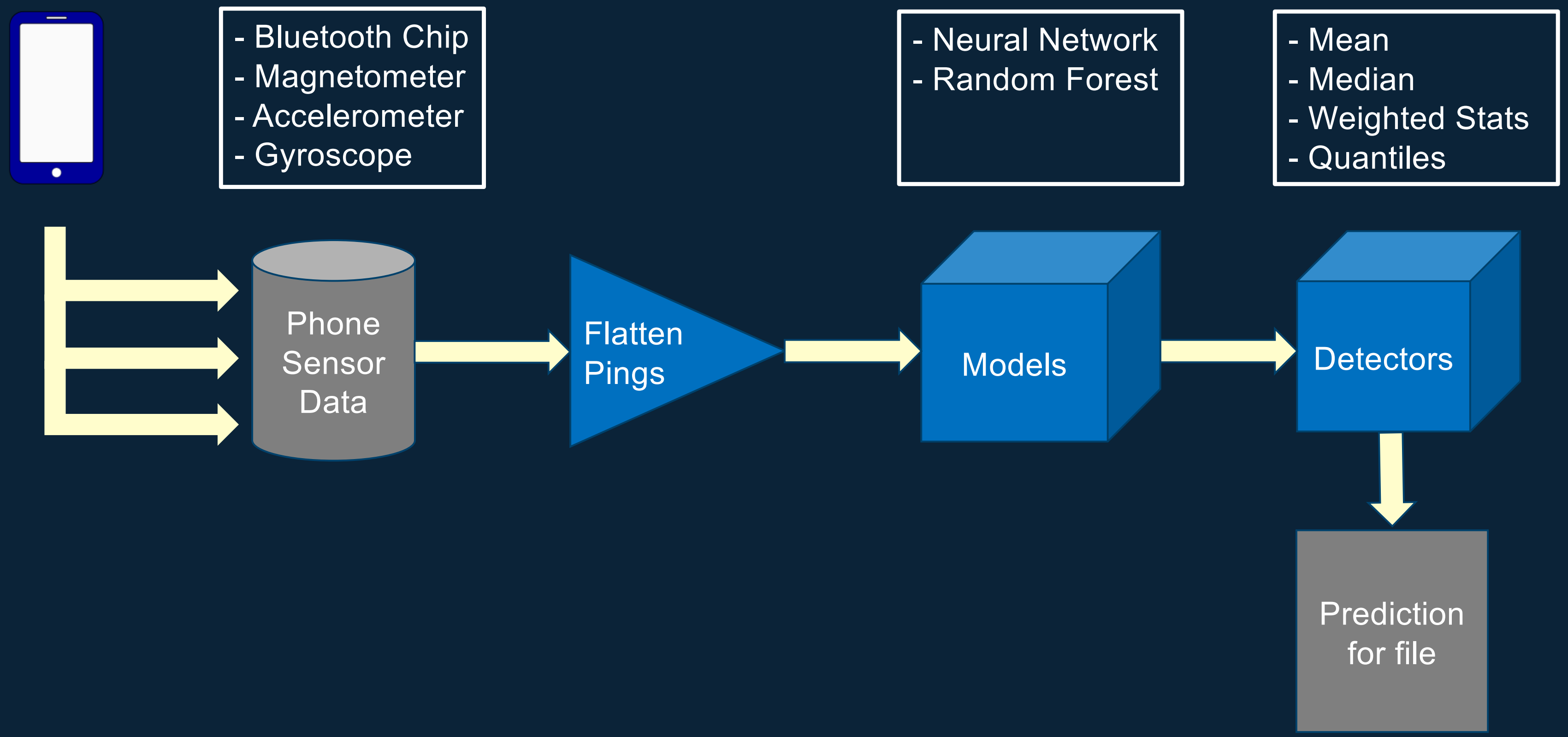
2 days of mlflow iteration



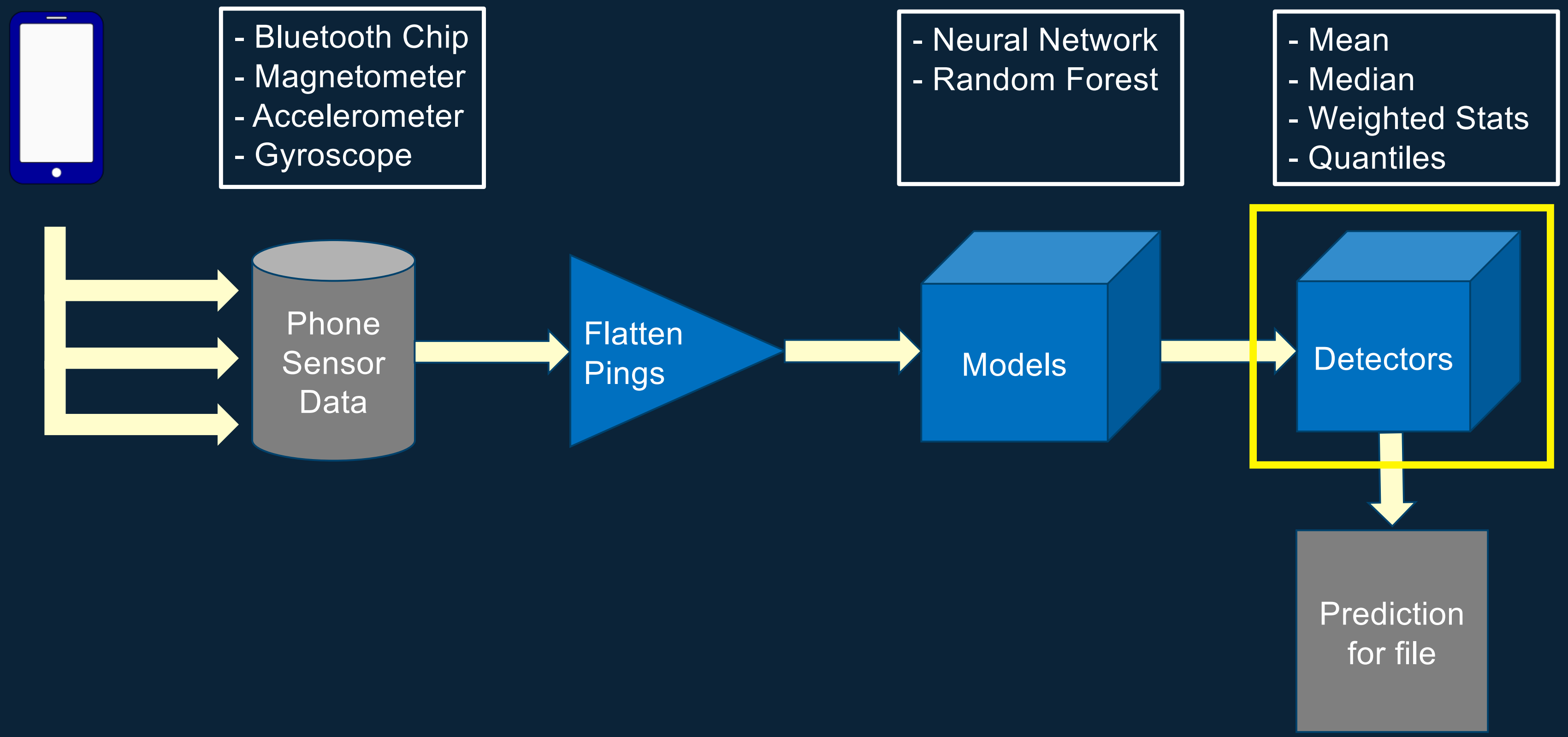
SUBSET	D	P_MISS	P_FA	NDCF
fine_grain	1.20	0.94	0.01	0.94
fine_grain	1.80	0.42	0.26	0.68
fine_grain	3.00	0.02	0.88	0.90
coarse_grain	1.80	0.35	0.25	0.60

MITRE

Pipeline Workflow



Pipeline Workflow



- Bluetooth Chip
- Magnetometer
- Accelerometer
- Gyroscope

Phone Sensor Data

Flatten Pings

- Neural Network
- Random Forest

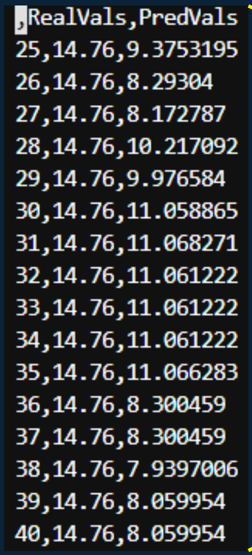
Models

- Mean
- Median
- Weighted Stats
- Quantiles

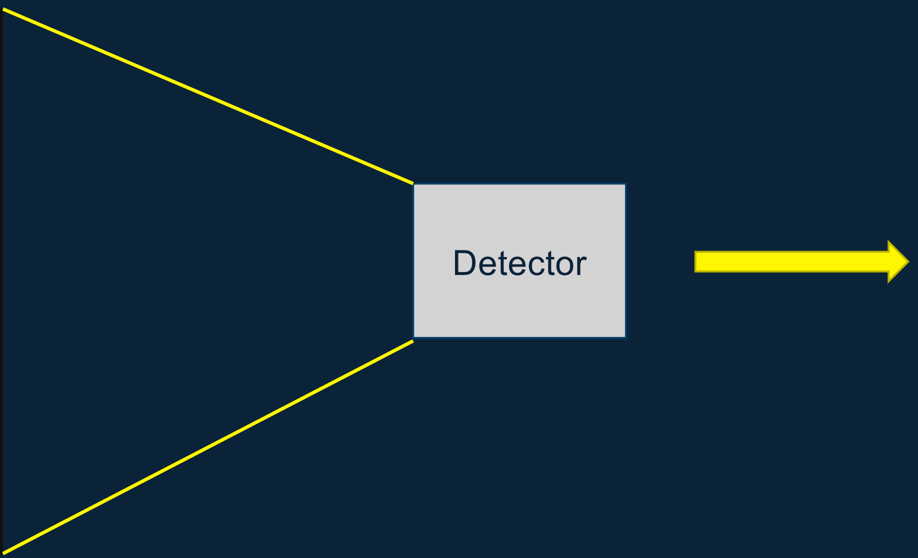
Detectors

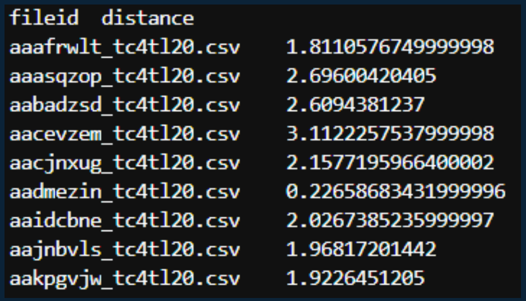
Prediction for file

Detector Summary



```
,RealVals,PredVals
25,14.76,9.3753195
26,14.76,8.29304
27,14.76,8.172787
28,14.76,10.217092
29,14.76,9.976584
30,14.76,11.058865
31,14.76,11.068271
32,14.76,11.061222
33,14.76,11.061222
34,14.76,11.061222
35,14.76,11.066283
36,14.76,8.300459
37,14.76,8.300459
38,14.76,7.9397006
39,14.76,8.059954
40,14.76,8.059954
```





```
fileid  distance
aaafrwlt_tc4tl20.csv    1.8110576749999998
aaasqzop_tc4tl20.csv    2.69600420405
aabadzsd_tc4tl20.csv    2.6094381237
aacevzem_tc4tl20.csv    3.1122257537999998
aacjnxug_tc4tl20.csv    2.1577195966400002
aadmezin_tc4tl20.csv    0.22658683431999996
aaidcbne_tc4tl20.csv    2.0267385235999997
aajnbvls_tc4tl20.csv    1.96817201442
aakpgvjw_tc4tl20.csv    1.9226451205
```

Detectors

Mean

- Simple mean over data file.
- Easy to implement.
- Prone to issues with outliers.

Median

- Similar to taking average.
- Less prone to outliers.

Weighted Average

- Can weight different output values differently.
- Weight values within 6 feet higher.
- Way to skew model outputs towards more “correct” values.

Model Summary

```
Model: "sequential“

Layer (type)              Output Shape              Param #
===========================================================
flatten (Flatten)         (None, 12)              0

dense (Dense)               (None, 16)              208

dense_1 (Dense)               (None, 16)              272

dense_2 (Dense)               (None, 1)              17
===========================================================
Total params: 497
Trainable params: 497
Non-trainable params: 0
```

Input Data

Magnetometer
Accelerometer
Gyroscope
Bluetooth RSSI
TX Power Level

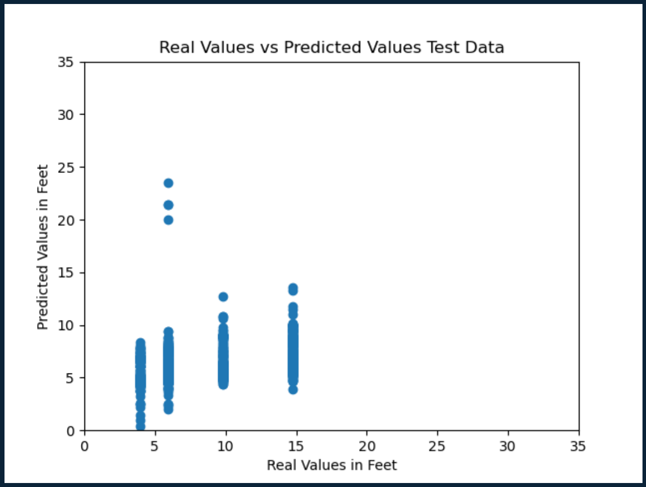
Training Parameters

Epochs: 15
Learning Rate: 0.001
Batch Size: 1500
Train/Test Split: 15/85

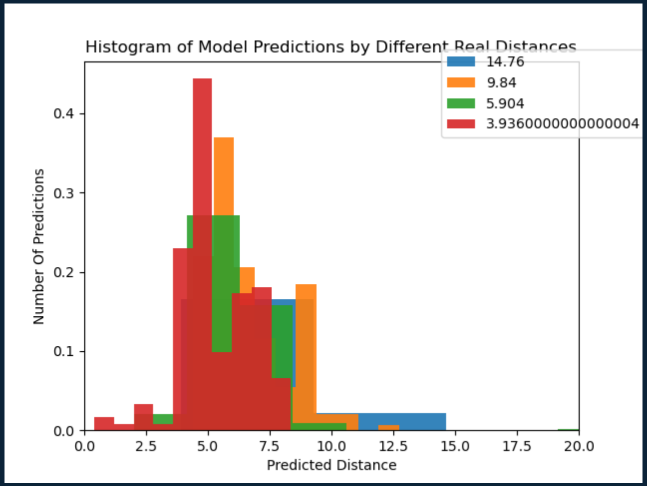
Optimizer: RMSProp

Training Time: 4.7 minutes

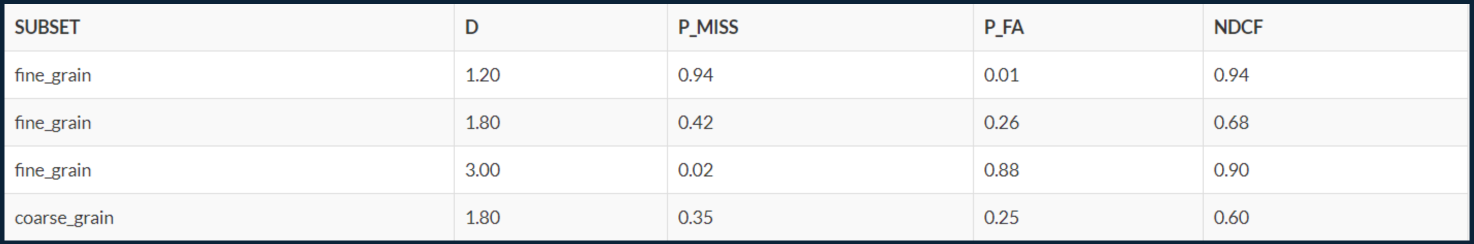
Model Result



Scatter Plot of Real vs Predicted values on Validation Set.



Density Histogram of Pings for Predicted Values by Real Distances for Validation set.



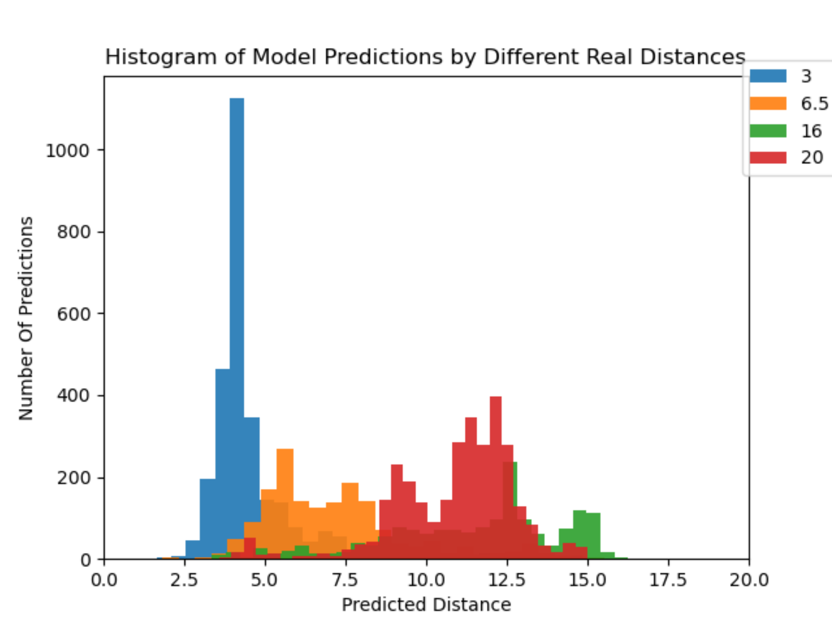
SUBSET	D	P_MISS	P_FA	NDCF
fine_grain	1.20	0.94	0.01	0.94
fine_grain	1.80	0.42	0.26	0.68
fine_grain	3.00	0.02	0.88	0.90
coarse_grain	1.80	0.35	0.25	0.60

Model Trained On Real World Data

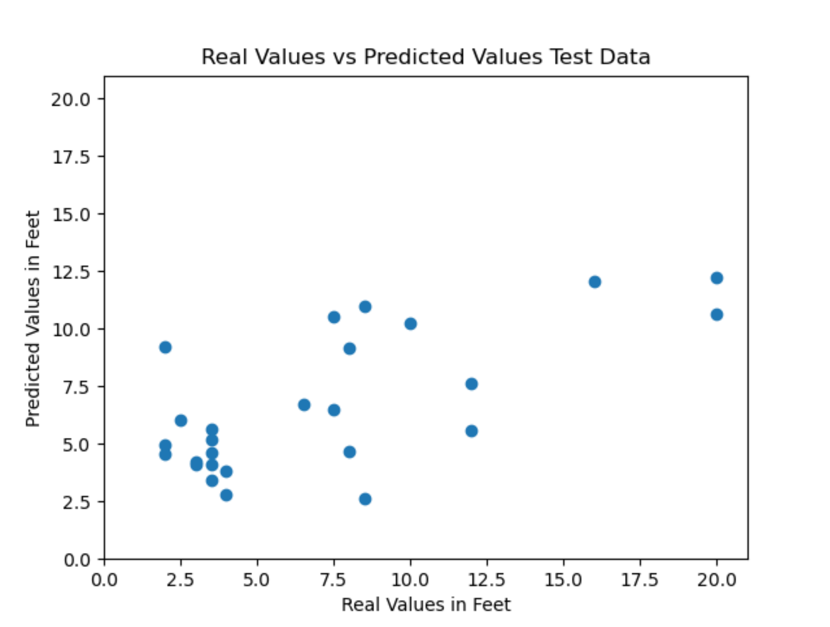
- Collected data personally on two iPhone 11’s.
- Data collected was from natural behavior at varying distances.
- Trained model on MITRE Range Angle Structured set and part of real-world data.
- Test data was never seen by model.

Idea to see how model would perform if training data was more realistic towards real world movement.

Neural Network Results on Real World Data



Histogram of model predictions after Neural Network model is run over test data.



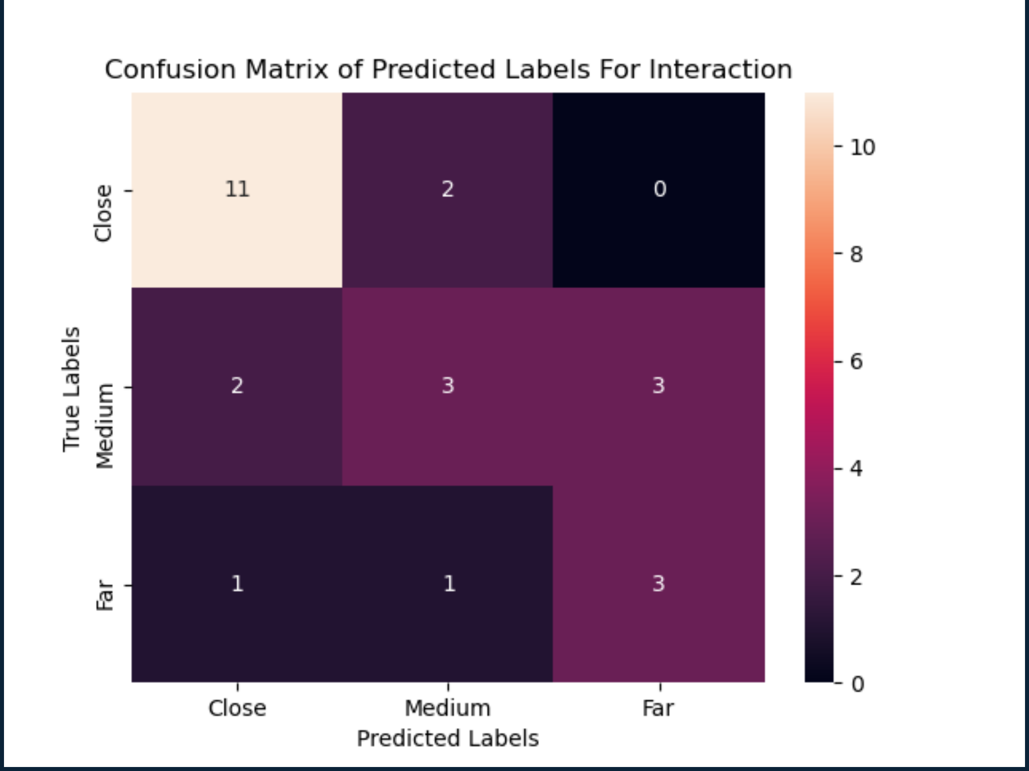
Scatter plot of real values versus predicted after median detector is run over file.

Neural Network Classification on Real World Data

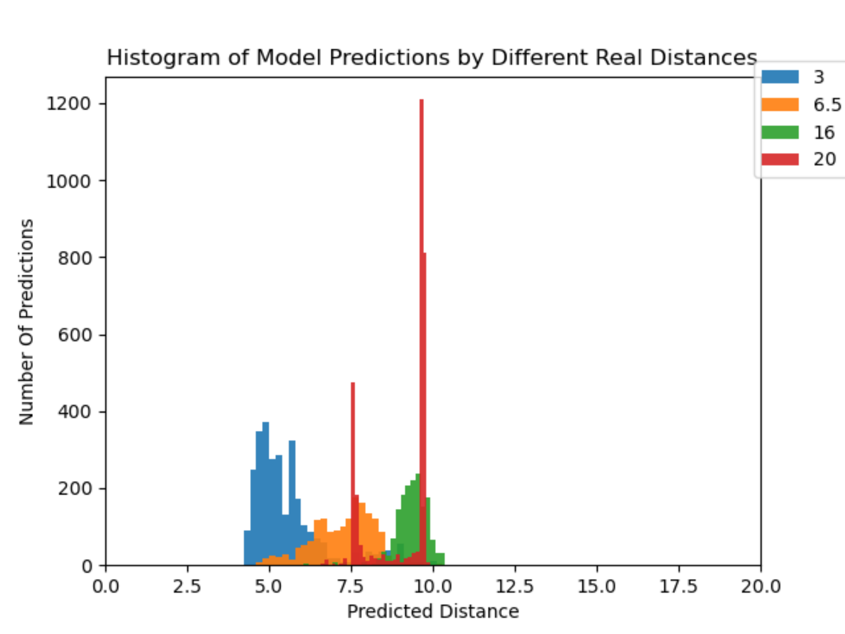
False Positive Rate: 3.8%

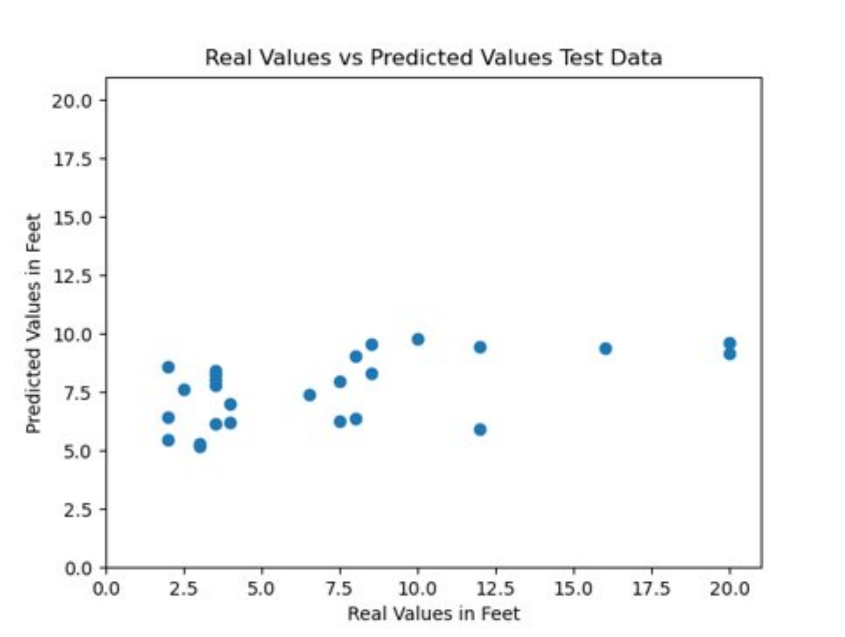
False Negative Rate: 7.7 %

Overall Accuracy: 65%



Random Forest





Histogram of model predictions after random forests model is run over test data.

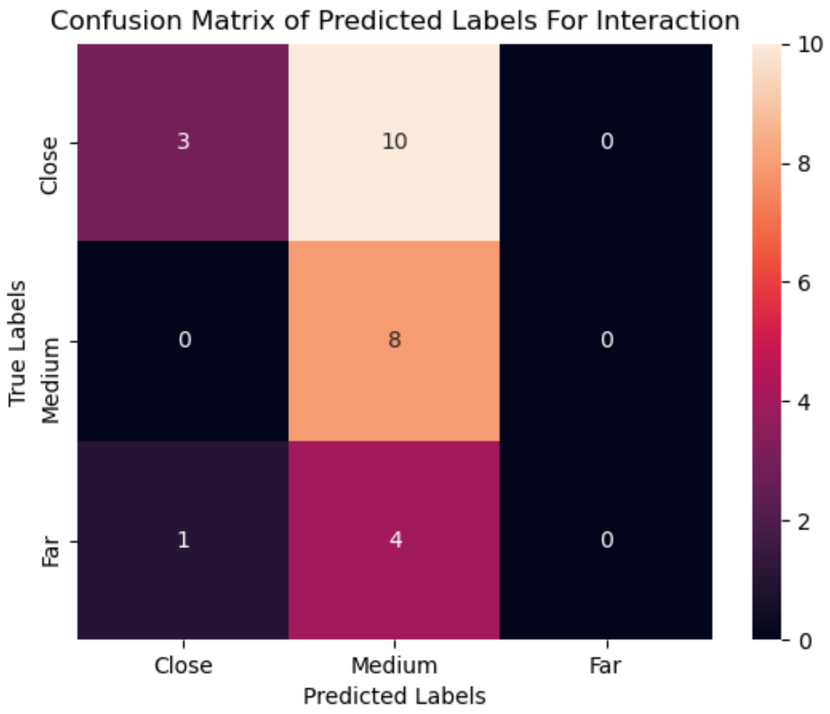
Scatter plot of real values versus predicted after median detector is run over file.

Random Forest Classification on Real World Data

False Positive Rate: 3.8%

False Negative Rate: 38.5%

Overall Accuracy: 42.3%



Confusion Matrix of Predicted Labels For Interaction

Lessons Learned

- Importance of validation data.
  - After we started including the NIST data for validation of our models our results improved substantially.
  - Was able to accurately see problem with model (overfit to data).
- Know the timeline.
  - After workflow was established to short a timeframe to make it fully effective.
  - Knowing the timeline would’ve allowed for more varied prototyping/exploration.

Next Steps

- More data augmentation/noise.
  - With augmentation to data would stop network from memorizing values.
- Implement callback to prevent overfitting.
  - Callback with early stop to prevent more overfitting.
- Additional data collection.
  - Collect more data from people behaving in a natural manner.
  - Hope is that by feeding in data with natural variability/close to what we would be getting the model will get more accurate.
- Work with multiple phones/environments.
  - Increase the scope of the model's accuracy by testing through different environments.
- Change model layout.